

Psychological

Monographs

General and Applied

p. 477
1959

Silverman

**A Q-Sort Study of the Validity of
Evaluations Made From
Projective Techniques**

By

Lloyd H. Silverman

New York University

Price \$1.00

Vol. 73
No. 7



Edited by Norman L. Munn

Published by the American Psychological Association, Inc.

Psychological Monographs: General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

NORMAN L. MUNN, Editor

Department of Psychology, Bowdoin College
Brunswick, Maine

Consulting Editors

ANNE ANASTASI
FRANK A. BEACH
ARNOLD M. BINDER
W. J. BROGDEN
ROBERT R. BUSH
JOHN F. DASHIELL
JAMES J. GIBSON
D. O. HEBB
EDNA HEIDBREDER
FRANCIS W. IRWIN

JAMES J. JENKINS
HAROLD E. JONES
DANIEL KATZ
BOYD McCANDLESS
DONALD W. MACKINNON
QUINN McNEMAR
LORRIN A. RIGGS
CARL R. ROGERS
RICHARD L. SOLOMON
ROSS STAGNER

Manuscripts and correspondence on editorial matters should be sent to the Editor. *Psychological Monographs* publishes comprehensive experimental investigations and programmatic studies which do not lend themselves to adequate presentation as journal articles. Major space is given to the author's original contribution; introductory and bibliographic materials, as well as statistical tables and graphs, must be kept within reasonable bounds. Tables, graphs, and appendix materials which deal with detail not essential to adequate presentation of the findings may be made available through the American Documentation Institute—for details of this procedure, see the *APA Publication Manual*. Preparation of manuscripts for publication as monographs should follow the procedure given in the *APA Publication Manual*. Publication in *Psychological Monographs* is free of cost to the author, except in cases where early publication is requested or author's alterations are made in galley proofs.

ARTHUR C. HOFFMAN, Managing Ed.; HELEN ORR, Promotion Mgr.; SADIE J. DOYLE, Editorial Asst.

Correspondence on business matters should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St., N.W., Washington 6, D.C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

COPYRIGHT, 1959, BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

A Q-SORT STUDY OF THE VALIDITY OF EVALUATIONS MADE FROM PROJECTIVE TECHNIQUES¹

LLOYD H. SILVERMAN

New York University

INTRODUCTION AND HISTORY OF THE PROBLEM

THIS STUDY was designed to investigate the validity of holistic projective technique evaluations. The holistic method treats the tests and the clinician as an inseparable combination. The clinician's appraisal is based on his total impression of the patient's productions and this appraisal is the unit evaluated. The question of the particular signs or content employed by the clinician in forming his appraisal is beyond the boundaries of this kind of research.

Of the previous research on holistic methods, the matching studies are the most numerous (Krugman, 1942; Palmer, 1951; Waehner, 1942). In these, judges are asked to match personality sketches drawn from an analysis of some projective test with sketches obtained from some other source. Investigations of this kind are inadequate in a number of ways. Since general sketches are usually used it is not possible to tell in which areas the projective technique in

question can evaluate well and in which areas poorly. A more crucial objection is that results become partially dependent on artifacts. Both the ability of the judges to match and the similarity in description of the reports to be matched become crucial determinants of the results. Spurious factors such as these can be responsible for the wide range of correlations found for different studies of this kind. Moreover, as Cronbach (1949) points out, the matching technique does not indicate the degree of rightness and wrongness for each prediction. Thus, a particular match or mismatch may be determined by the smallest of coincidences.

Some of the studies approaching projective technique validation from a holistic point of view avoided the pitfalls of the matching method by restricting the psychologists' task to designating a diagnostic category for the patient. A number of these have demonstrated a high degree of agreement between the psychologists' diagnoses and the criterion (Benjamin & Ebaugh, 1938; Chamber & Hamlin, 1957; Siegel,

¹ This monograph is based upon a doctoral dissertation at New York University in which the research is reported in much greater detail (Silverman, 1958). The study was carried out at the Psychiatric Clinic of the Court of Special Sessions of the City of New York. I wish to express my deepest appreciation to Jules S. Golden, director of the clinic, who graciously permitted me to conduct this research and without whose cooperation the procedures employed in this investigation would not have been possible. Thanks are also due to the ten staff psychiatrists who served as psychiatric judges. I also wish to express my gratitude to the following thirty clinical psychologists who evaluated the projective test material, kindly giving time and effort without compensation: Stanley Berger, Fred Brown, Renata Calabresi, Louis Feigenbaum, Flor-

ence Halpern, Howard Halpern, Emanuel Hammer, Doris Heller, Robert Holt, Walter Kass, Gertrude Kurth, Frank Lachmann, Leah Levinger, Dorothy Litwin, Carola Mann, David Mann, Ruth Munroe, Martha Schon, Miriam Siegel, Herbert Spohn, Bernard Steinzor, Allen Williams, Berta Beller, Lawrence Epstein, Ladilly Harris, Leone Lesser, Adam Munz, Irving Schwartz, Irving Steingart, and Joan Trachtman. Acknowledgment is also due to Fred Brown, chairman, Robert Holt, and Isidor Chein, who as members of the doctoral committee gave guidance and encouragement throughout this study, and to Jacob Cohen who graciously offered me advice and assistance with respect to the statistical procedures employed.

The author is also Senior Psychologist at Mt. Sinai Hospital, New York City.

1948). Investigations of this kind are important inasmuch as they show in an objective scientific manner the validity of projective techniques in one particular area. However, the worth of projective techniques is purported to extend much further than merely placing an individual into a category of diagnosis. Clinical psychologists have long maintained that the richness of projective methods lies in their ability to discover the more refined and less obvious aspects of personality. Schafer (1954) has stressed the great value of the Rorschach in discerning defense mechanisms. Brown² believes that much regarding personality dynamics and the person's earliest perceptions of significant figures can be elicited from projective material. Piotrowski³ emphasizes that self concept and role in life are frequently revealed in Rorschach responses.

A few attempts have been made to validate projective techniques in terms of the variables mentioned above. Typical of these is the famous one by Hertz and Rubinstein (1939). Here, a Rorschach was given to one *S* and interpreted "blindly" by three leading Rorschachers (Hertz, Beck, and Klopfer). The interpretations were validated against clinical material gathered in 14 interviews. From an over-all perusal of the four reports, a great deal of similarity was said to exist both among the Rorschach experts and between each of them and the validating criterion. Studies such as this, although dealing with more extensive material than just a diagnostic classification, leave much to be desired as validation attempts. The fact that a single case is usually used is an obvious shortcoming. Even more important is the fact that the method used for evaluation does not allow for a systematic comparison of the psychiatric and psychological evaluations. Moreover, in the absence of a numerical index, the amount of agreement between the psychologists and the criterion can only be subjectively judged by the individual reader. This task

is sometimes made especially difficult by the fact that the psychotherapists and the projective test evaluators discuss different aspects of the same patient.

Krugman (1942) tried to overcome these shortcomings by asking judges to rate on a four-point scale the degree of similarity between Rorschach reports and case study abstracts for different areas of functioning. In 94% of the cases there was "essential" or "fair" agreement. Similarly, Symonds (1955) found average agreement at 65% when he himself used a scale to judge the degree of similarity between Rorschach reports written by seven experienced Rorschachers and extensive case study material on a single patient. Although an improvement over the aforementioned studies, these still did not overcome all of the objections previously raised. The case study abstracts still contained material for areas the Rorschach report did not deal with and vice versa; and in terms of subjectivity in evaluating the degree of agreement, all that was accomplished was that the opinions of judges were substituted for the opinions of the reader. Hertz's 1941 characterization of the case study method still aptly applied: "The case study approach, though quite fruitful, remains a clinical qualitative approach rather than a quantitative scientific one. It may be that new statistical devices will be created that will be able to place this approach through a quantitative analysis of the clinical material itself in a position of good validation" (Hertz, 1941, p. 531).

One statistical device that has been so employed is the rating scale (Samuels, 1952; Saxe, 1950). A second quantitative technique that has been used is the *Q* sort. This is superior to usual rating scales since it asks the evaluator to judge items in a relative sense (that is, the relevance of one item compared with others), while the rating scales demand that items be judged in absolute terms. The latter can well lead to spuriously low agreement since different raters, whose evaluations are to be correlated with each other, may adopt different anchoring points for their ratings.

In light of this advantage of *Q* sorts, it is surprising that only two studies could be

² Brown, F. Personal communication, 1958.

³ Piotrowski, Z. Personal communication, 1955.

found in the literature that used them in validating projective techniques. One was reported by Little and Shneidman (1955). This was a quantitative addendum to Shneidman's (1951) Thematic Test Analysis book which presented in report form appraisals by 17 "expert" psychologists of one patient's TAT and MAPS protocols. Statements were culled from these reports and items stating their converse were formulated. One hundred fifty of these items were chosen for quantitative assessment. These statements were resubmitted to the 17 psychologists for *Q* sorting. The criterion measures were the sortings of the statements by 29 "competent clinicians" who independently appraised the statements on the basis of a complete clinical record (therapy notes, hospital observations, etc.). Validity coefficients between each psychologist in the first group with every clinician in the second group ranged from $-.07$ to $.70$, with the mean at $.45$.

The second study using *Q* sorts was reported by Fisher (1952). She compared the *Q* sorts of eight "expert" clinical psychologists with those of four psychiatrists for five patients, each of whom was receiving psychotherapy from one of the psychiatrists. Four of the psychologists based their evaluations on figure drawings of the *Ss* while the other four had accessible instead a Rorschach, a TAT, and a Stanford-Binet intelligence test. The over-all correlation between the first group of psychologists and the criterion was $.195$, while the correlation between the second group and the criterion was $.365$. This difference was ascribed to the greater number of tests the second group had available.

Despite their use of *Q* sorts, these studies were unsatisfactory in two major respects. First, in both studies only one *Q* sort was utilized which included items on different levels of personality functioning. Thus, some of them were diagnostic statements, others referred to unconscious needs, still others were concerned with conscious feelings, and some even referred to overt behavior. By having statements at so many different levels in the same *Q* sort, spurious differences between evaluators were apt to

arise. One evaluator, because of personal preference, theoretical orientation, or the particular medium he works with, may give priority to statements on one level over those on another. Thus, differences between sorters were apt to arise due to artifacts rather than real disagreement. That this actually occurred is implicit in Little and Shneidman's findings that statistical analysis revealed that their psychologists were composed of two groups: one of which preferred to deal with dynamic inferences, and the other with inferences related to the kind of psychopathology present.

Secondly, in neither study were controls utilized. That is, there was no way of telling to what degree the *Q* sorts were "stacked" in that some of the items may have applied to all or most patients in a particular age group while others may have had little or no applicability. The likelihood that this actually occurred at least in Little and Shneidman's study is strongly suggested by an examination of the 24 items these authors reported as being ranked highest and lowest on the *Q*-sort continuum. All 12 of the items ranked highest in applicability dealt with pathological aspects of the patient (for example, "he is in the early stages of paranoid schizophrenia"). Conversely, all 12 of the items ranked lowest had positive connotations ("he is not a particularly sick individual"). One could legitimately ask whether the psychologists could not have attained almost as high correlations if they based their rankings on just the knowledge that the *S* was a patient, without any test data available.

In addition to these two failings of the studies by Fisher and by Little and Shneidman, two more limitations restricted the applicability of their results. These were: (a) only clinical psychologists with a great deal of experience and reputations as "experts" were employed in both pieces of research. Thus, no evidence was forthcoming as to the degree of predictive accuracy of clinicians of less experience and reputation. (b) Since only one heterogeneous *Q* sort was used, only an over-all statement could be made as to the effectiveness of psychologists' appraising personality via

projective methods. No data were made available as to their relative effectiveness for determining dynamics, defense mechanisms, overt behavior, diagnosis, etc. In the present investigation, an attempt will be made to overcome the shortcomings and limitations of these two pioneer attempts.

HYPOTHESES

Hypothesis I. There is a positive relationship exceeding chance agreement between the evaluations of clinical psychologists working with the projective test material of patients and the evaluations of psychiatrists working with the same patients in psychotherapy.

A number of prior studies investigating the validity of projective techniques have reported a great deal of variability in the ability of different clinical psychologists who acted as participants (Chamber & Hamlin, 1957; Samuels, 1952; Symonds, 1955). Yet there has been a lack of investigation as to what the crucial differences are between those clinicians whose evaluations correlate well with the criterion measure and those whose correlations do not. This study will investigate differences in terms of two variables that should be pertinent in this regard. Hypothesis II is concerned with the experience variable. The expectation of differences here is based on the principle that in any endeavor requiring complex skills, performance will improve with experience.

Hypothesis II. The degree of positive relationship between psychologists' and psychiatrists' evaluations will be, in part, a function of how experienced the psychologist is; the greater his experience, the greater his accuracy in appraising the patient.

Hypothesis III is concerned with the personal analysis variable. The expectation that there will be a difference in the ability of those psychologists who have undergone Freudian psychoanalysis and those psychologists who have not, in making projective test evaluations, is based on the following rationale. The results of a number of studies have demonstrated that a positive

relationship exists between lack of insight into self and lack of insight into others (Frenkel-Brunswick, 1951; Goodman, 1952; Norman, 1953; Sears, 1936). Since projective test evaluations involve gaining insight into others, it would follow that the degree of accuracy with which psychologists appraise a patient will, in part, depend on the degree of insight that they have into themselves. Experimental findings supporting this idea have, in fact, been reported by Mintz (1955) and Filer (1951). Since the therapeutic process in Freudian psychoanalysis centers upon the development of self-insight (Bibring, 1954), psychologists who have received such treatment would be expected to show greater accuracy in their appraisals.

During the course of a Freudian psychoanalysis, self-insight is developed at many different levels. According to Bibring, these include the patient's understanding that he reacts in typical ways to typical situations; that certain of his attitudes, which at first appeared to him unrelated, are in fact related to each other; that certain reaction patterns form a characteristic sequence; and that he is motivated by specific unconscious ideas and feelings. Freudian psychoanalysis aims at an understanding of all of these, but the understanding of unconscious material plays the central role (Bibring, 1954). Thus, a clinical psychologist who has undergone such treatment should be at a particular advantage in his projective test evaluations when unconscious forces of the patient need to be appraised. Especially sensitive to these forces in himself, he would be expected to be especially sensitive to these forces in others. Since some of the areas to be evaluated in this research are concerned with unconscious aspects of personality, there is thus an added reason for anticipating differences in terms of the personal analysis variable.

Hypothesis III. The degree of positive relationship between psychologists' and psychiatrists' evaluations will be greater for those psychologists who have undergone Freudian psychoanalysis than for those who have not. The reason for not hypothesizing as great a degree of positive relationship

for those psychologists who have received other forms of treatment as for those who have undergone Freudian psychoanalysis rests on my taking seriously a view of Bibring and thus my wishing to test one implication of this view. Bibring writes that one of the differences between Freudian psychoanalysis and the various types of dynamic psychotherapies derived from it is that in the latter there is "a general trend to shift the emphasis from insight through interpretation toward 'experimental' manipulation, that is learning from experience seems to become the supreme agent rather than insight through interpretation" (Bibring, 1954, p. 766).

Hypothesis IV is concerned with the increasing agreement between the psychiatrists' successive evaluations of his patient and the evaluations of the clinical psychologists. This bears upon the practical value of having patients who are entering psychotherapy tested beforehand.

I could find only one study in the literature bearing on this issue. Siegel (1948) compared the degree of agreement between two psychiatric diagnoses, made one year apart, and the diagnosis made on the basis of a Rorschach. She found 88.5% agreement between the Rorschach and the second psychiatric diagnosis, compared to only 61.5% agreement between the Rorschach and the first psychiatric diagnosis. Only one psychologist participated in the study and only the area of diagnosis was investigated so that further study of this question is very much in order.

Hypothesis IV. The degree of positive relationship between the psychologists' and psychiatrists' evaluations will be, in part, a function of the length of time the psychiatrist has seen the patient. There will be an increasing correlation between the psychiatrists' successive evaluations of their patients and the evaluations of the clinical psychologists. This is equivalent to saying that, as the psychiatrists' behavior samples are enlarged, their judgments become more valid, and, as the psychiatrists' judgments become more valid, they converge upon the test evaluations.

Hypotheses V and VI are concerned with interaction effects between the variable "length of time the psychiatrist has seen the patient before making the evaluation" and the variables of experience and personal analysis. The expectation that the latter two variables will affect the degree to which psychologists can discern the hidden core of personality, thus perceiving aspects of the patient that are not clinically apparent for some time, is based on essentially the same rationales as those underlying Hypothesis II and Hypothesis III.

Hypothesis V. The greater the experience of the psychologist, the more pronounced will be the increasing correlation between his evaluation of the patient and the psychiatrist's successive evaluations of the patient.

Hypothesis VI. There will be a more pronounced increasing correlation between judgments by those psychologists who have undergone Freudian psychoanalysis and the psychiatrists' successive evaluations than between judgments by those psychologists who have not undergone such treatment and the psychiatrists' successive evaluations.

METHODS USED

Selection of Cases

The Ss whose projective test material was to be evaluated were 10 young adult males between the ages of 17 and 22 who committed crimes that brought them to the Court of Special Sessions of the City of New York. From here, they were either referred directly by the judge or through the probation department to the psychiatric clinic attached to the court. They were chosen from among the total patient population of the clinic in the following manner. As soon as it was decided to embark upon this project, each new patient seen at the clinic who was judged suitable for once- or twice-a-week psychotherapy was considered a potential research case. This meant that he had to have administered to him the various projective techniques that were to make up the research test battery prior to his beginning treatment. It was decided that for the purpose of consistency, I should administer all the projective tests for potential research cases.

This procedure was followed for many more than 10 Ss since it was anticipated that some of the Ss would not remain in therapy long enough so that an extended psychiatric appraisal could be made. Thirty-five was decided upon as the minimum

number of therapy sessions that had to be completed for the patient to be considered a research case. This caution turned out to be fully justified inasmuch as about 25 persons had to be considered as research potentials before 10 could be found who completed the minimum number of psychiatric sessions. The others either refused to continue treatment at some point before 35 sessions or else were rearrested and given jail terms so that therapy could not continue.

The Projective Test Battery

The projective test battery administered to the research cases consisted of the following techniques:

Rorschach. Administered in accordance with the Klopfer technique (Klopfer & Kelley, 1942).

Thematic Apperception Test. Administered in a manner similar to that outlined by Murray (1943) with the following exceptions: (a) There was no time limit imposed for each story. (b) Only one testing session was held, this following the procedure described in Murray's "first session." (c) The following ten cards were presented to each S: 1, 2, 6 BM, 7 BM, 7 GF, 8 BM, 9 GF, 12 M, 13 MF, and 18 GF. In addition, on the basis of my clinical judgment regarding what other cards would yield significant material in that particular case, between two and nine additional stories were obtained from each S from among cards 3 BM, 4, 5, 6 GF, 9 BM, 10, 13 B, 17 BM, and 18 BM.

House-Tree-Person Drawings. Administered in accordance with the Buck (1948) technique except for the fact that regular 8½ × 11-in. paper was used. In addition, inquiry was made about the figures drawn and, where it seemed appropriate, about the house and the tree as well.

Most-Unpleasant-Concept Test. Administered in accordance with the Harrower (1952) technique.

The first three of these projective tests were chosen because of their very wide use in clinical practice. The fourth (Most-Unpleasant-Concept) was included because of the very small amount of time needed for its administration which I have found to be amply justified by the richness of the data it yields.

Clinical Psychologist Evaluators

Thirty clinical psychologists took part in evaluating the projective test protocols. They were divided into groups in terms of two variables:

Experience variable—Group I. Ten clinical psychologists who have had 10 or more years of projective testing experience and who have earned reputations through their teaching, supervising, and/or writing for their high competence in dealing with projective test material. The decision as to whether the psychologist had earned such a reputation was made by me with the concurrence of at least one member of my thesis committee. *Group II.* Ten clinical psychologists who have had

between five and eight years of projective testing experience. *Group III.* Ten clinical psychologists who have had three or fewer years of projective testing experience, including their internship.

Personal analysis variable—Group A. Ten clinical psychologists who either completed a Freudian psychoanalysis or who completed a large part of their analysis although still in treatment. This group hereafter will be referred to as the "psychoanalyzed group," and its members as the "psychoanalyzed psychologists."⁴ *Group B.* Nineteen clinical psychologists who either had not undergone Freudian psychoanalysis at all, or, as was the case with three participants, had recently begun such treatment but had not completed more than four months of it. This group hereafter will be referred to as the "nonpsychoanalyzed group," and its members as the "nonpsychoanalyzed psychologists." One psychologist, who had been in Freudian psychoanalysis for one year and then had left, could not qualify as belonging to either of the two groups. Thus, her evaluations were not included when the variable of personal analysis was considered.

It was noted that for further comparisons, the psychologists in Group B could be subdivided into: Subgroup B₁—nine psychologists who had either completed a course of individual treatment other than Freudian psychoanalysis or who had completed a large part of such treatment. The members of this group varied in terms of the orientation of their therapists and the frequency of their sessions. Some described their treatment as once- or twice-a-week psychotherapy and others as Sullivanian analysis. For convenience, this group hereafter will be referred to as the "psychotherapy group."⁵ Subgroup B₂—nine psychologists, among whom seven had received no amount of individual psychotherapy and two who had just begun a Freudian psychoanalysis. This group hereafter will be referred to as the "nontreated group." One of the nineteen psychologists in Group B had received some form of treatment other than Freudian psychoanalysis for a period of between six and sixteen months during the course of his participation in the research. This was thought to be too short a period of time for him to qualify for Subgroup B₁ and too long a period of time for him to qualify

⁴ Five members of this group had completed a Freudian psychoanalysis, two were in the sixth year of their analyses, one in his fifth year, and two in their third year.

⁵ Five members of this group had completed their treatment and four members had completed at least three years of treatment. In terms of the orientation of their therapists, six of the psychologists had received treatment from a member of the interpersonal school, one received treatment from a therapist who was described as a follower of Stekel, and two who described their therapists as eclectic in orientation.

for Subgroup B₁. Thus, when these two subgroups were compared and when a three-way comparison was made, the evaluations of this psychologist were not included. An approximately equal proportion of participants from each of the three experience groups was included in Groups A and B on the one hand, and Subgroups B₁ and B₂ on the other. This meant that when either the experience variable or the personal analysis variable was to be considered separately, the other was held constant.

Distribution of Cases Among Psychologists

Each psychologist was asked to evaluate the protocols for two of the ten patients. Thus, 60 psychological evaluations were made. In order to keep equal the number of psychologists doing each case, the protocols for each patient were evaluated by six psychologists.

In order for the most meaningful comparisons to be made in regard to differences between psychologists at the three experience levels, each case was assigned to two psychologists from Group I, two psychologists from Group II, and two psychologists from Group III. In regard to the personal analysis variable, at least one psychologist representing Group A and Group B and at least one psychologist from Subgroup B₁ and Subgroup B₂ participated in each case.

For appraising each case the psychologist was provided with a copy of the patient's Rorschach, TAT, and drawings of house, tree, persons, and most unpleasant concept. In addition, the psychologist was informed of the patient's age, sex, race, nationality, the number and sex of siblings, and whether or not the patient was currently living at home. A brief statement was also made about the patient's upbringing in terms of who reared him (one or both parents, or some other person). The psychologists were also told that all the cases came from the files of a psychiatric clinic attached to the court. Under ideal conditions, the latter fact should not have been made known to the psychologists in a study such as this. However, some of the participating psychologists who were closely associated with me knew the source from which the cases were being drawn. Thus, to equalize the conditions under which the various psychologists operated, this fact was revealed to all, and a control had to be instituted (to be described later) to determine if this knowledge influenced the results.

Criteria for Validation

The degree of validity of the psychologists' evaluations was determined by correlating them with evaluations made by the psychiatrist treating the patient. In addition to his contact with the patient in therapy, the psychiatrist in making his evaluation was free to utilize material gathered by the clinic social worker who had interviews with the parents of the patients, the report of the in-

take psychiatrist, and the probation report on the patient. In no instance, however, were the results of the psychological tests given at intake made available to the psychiatrist. This criterion evaluation took place at a point after 35 or more therapy sessions.

Background of the Participating Psychiatrists

Since each patient was seen by a different psychiatrist, 10 psychiatrists participated. They had had between five and ten years of psychiatric experience either as residents, in clinical appointments, or in private practice. The mean number of years of experience was eight. All the psychiatrists had either completed postresidency psychotherapeutic or psychoanalytic training or had had at least one and one-half years of such training. The mean number of years completed was three. They had all received some form of psychotherapy, either having completed it, or having completed two and one-half years or more of it, though still in treatment.

The kind of training and personal psychotherapy that the psychiatrists received varied. Two of them were being trained in Freudian psychoanalysis, one at the New York Psychoanalytic Institute and one at the Psychoanalytic Institute of the State University of New York. Four psychiatrists had received training and therapy at the William Alanson White Institute, a training center of the interpersonal school. The remaining four psychiatrists had received training and therapy at the Psychoanalytic Clinic for Training and Research at Columbia University. This center was under the directorship of Sandor Rado during the time that the four psychiatrists were trained there and Rado's orientation is no longer considered Freudian psychoanalytic (Glover, 1957). Consistent with this view were my impressions gotten from frequent clinic staff conferences that were attended by all the psychiatrists. From the discussions of case material, it appeared that the approach of the psychiatrists from the Columbia school, both in terms of theory and therapy, was much closer to the approach of the psychiatrists trained at the William Alanson White Institute than those being trained in Freudian psychoanalysis.

Early Psychiatric Evaluations

The psychiatrists were also asked to evaluate the patients at two earlier intervals during treatment, after 4 to 8 sessions, and after 18 to 22 sessions. These evaluations were also correlated with the evaluations of the clinical psychologists, as was the criterion evaluation of the psychiatrists. A comparison of the three sets of correlations thus could be made to determine if there were differences in the amount of agreement between psychiatrists and psychologists at the three intervals.

Method of Evaluation

The method of evaluation for both psychologists and psychiatrists involved the use of *Q* sorts. The evaluator was given six separate sets of statements, one for each of six areas to be appraised. Each set consisted of 30 statements and was dealt with separately. The 30 statements were sorted into seven groups ranging from most to least applicable in accordance with the following distribution: one—1, three—2, six—3, ten—4, six—5, three—6, and one—7. The rationale for this procedure has been outlined by Stephenson (1953).

In the second and in the final evaluation by the psychiatrists, I reminded them to appraise the patient in terms of the type of person he was when the projective techniques were administered, excluding any changes that may have occurred as a result of psychotherapy.

In the appraisals by the psychologists, the projective test material was first left with them so that they could familiarize themselves with it and make whatever notes were necessary for the evaluation. They were asked to spend about as much time looking over the material as they would ordinarily devote to looking over case material in their clinical practice. Then when they met with me for the actual *Q*-sort evaluation, they did not have to spend too much time referring back to the raw data.

During the first and second evaluations by many of the psychiatrists, I noted that despite the fact that no limit was placed on the amount of time they could spend on making their evaluations and despite the fact that they were compensated for their time, they sometimes rushed through the procedure. At times, their decisions appeared to be made quickly and impulsively rather than as the result of reflective thought. Since the third and final evaluation of the psychiatrist was to be the criterion measure against which the psychologist's evaluations were going to be compared, it was felt that the psychiatrists should be encouraged to spend more time on this final evaluation. The following steps were therefore taken. First, the third evaluation was allowed to proceed in the same manner as the first two evaluations. Next, it was explained to the psychiatrist that further meetings with him would be necessary in order for him to check his third evaluation; that since this was to be the criterion evaluation, it should be as accurate as possible. Then, additional meetings were held in which an extended checking procedure took place for each of the six evaluation areas. For every psychiatrist, at least a few shifts in ratings were made in each area and it was my impression that these shifts were the consequence of greater care taken and more time spent in consideration.

Areas of Evaluation

Each time a patient was appraised by a psychologist or psychiatrist, evaluations were made

separately for six areas. The items utilized in the six *Q* sorts appear in the appendix. Each of the areas will now be discussed separately:

Defense mechanisms. The items in this area all dealt with the means by which people protect themselves against anxiety generated by intrapsychic conflict. To insure for uniformity in approach, all the evaluators were instructed to consider as the most applicable defense mechanisms those that were most characteristic of the particular patient rather than those that were used quantitatively the most often. The example of "repression" was given, this defense probably used quantitatively more often than other defenses by most people. However, this defense was not to be considered as one of the most applicable unless it especially characterized the particular patient as it would, for example, in the case of a classical hysteric.

Motivating needs and affects. Some of the items in this area pertained almost exclusively to unconscious motivating forces (for example, those referring to castration fears and incestuous longings). Other motivators, however, such as guilt over heterosexual impulses and exhibitionistic-voyeuristic needs, are conscious or preconscious in some persons and unconscious in others. In appraising this area, all evaluators were instructed to ignore the question of whether the motivator under consideration was conscious or unconscious to the patient but to consider only whether and to what degree it was motivating him. For example, the item on "feelings of inferiority" was meant to apply to a patient if he was either consciously plagued by feeling less able than other people or only aware of feelings of grandiosity which served as a compensatory defense against the inferiority feelings.

The decision to ignore the level of consciousness was made for two reasons. First, it would seem much more appropriate for projective testers to concern themselves with the degree to which a particular force was present rather than with whether it was conscious or not, since the latter question could be answered more readily by the psychotherapist. Secondly, from my contact with projective testing, it seems rather doubtful to me that the question of whether a patient's strivings are conscious or unconscious can be answered on the basis of projective tests alone.

In rating the items in this area, to insure for uniformity of purpose, the problem of level (i.e., immediacy) also had to be dealt with, for it could be justifiably maintained that most of these motivating forces play a significant role in everyone's personality development. Thus, the evaluators were told to give priority to those motivators that were influencing the patient at the present time; needs and fears that would reveal themselves in current dreams, fantasies, slips of the tongue, etc. Other motivating forces, even if judged to be at the core of the patient's illness, were to be considered as secondary if they were judged to

have no direct discernible influence on the patient's behavior at the present time.

Character traits. In evaluating this area, all evaluators were instructed that the 30 statements contained therein were all meant to apply to deeply embedded and long-lasting characteristics of the patient and not to transient or easily modifiable traits.

Diagnosis and symptoms. The statements in this area were of two types. Some dealt with character structure and others dealt with the symptomatic picture. The latter items were stated in terms of the patient's vulnerability to the particular symptom or syndrome. The rationale for this type of formulation is that it seems more useful and more appropriate for a projective tester to call attention to the patient's potential for developing "psychosomatic symptoms," "sexual perversion," "sexual impotency," "schizophrenia," "anxiety states," etc., than to decide whether or not such a state existed at the particular time the patient was tested. It is more useful in the sense that the therapist can usually quite easily discern for himself the presence of a symptom or syndrome. And it is more appropriate because very often the actual manifestation of these disorders depends on external circumstance or the patient's physical condition, in addition to his internal psychological state—factors the psychologist may not be aware of.

Interpersonal behavior. The question of whether statements about a patient's overt behavior with other persons, particularly with as much specificity as is demanded in the items in this *Q* sort, could be made from projective test material has been debated among clinical psychologists. It was because of the disagreement on this issue that this area was included in the hope that the results obtained might lend support to one or the other viewpoint. In appraising this area the psychologists and psychiatrists were asked to be guided only by how they thought the patient typically acted, disregarding entirely the question of motivation.

Infancy and childhood as perceived by the patient. This area, like the last mentioned, deals with material over which clinical psychologists are in disagreement as to its predictability from projective tests, a difference of opinion which the results of this study might bear upon. I emphasized to the evaluators when they were making their appraisals in this area that these items referred only to how the patient perceived parental figures during infancy and childhood and not to how the parental figures actually behaved. As in the area of motivating needs and affects, the evaluators were told that they did not have to concern themselves with the question of whether these perceptions were currently conscious to the patient or, for that matter, if they were ever conscious. The criterion by which their applicability was to be judged was the degree to which these perceptions played a role in influencing the patient's personality development.

Selection of Statements for the Q Sorts

The items used in the *Q* sorts were obtained in the following manner. Several hundred statements were extracted from psychological reports written by experienced clinical psychologists for psychotherapists. I then divided them into the six aforementioned areas. After this they were presented to a group of five psychologists who were asked to eliminate all items that they considered to be vague or ambiguous, inappropriate to the area to be evaluated, and repetitious. I made further eliminations and some additions and rephrased many of the items so that each of the *Q* sorts finally contained 30 statements, each of which I judged to be stated clearly and concisely.

Definition of Terms

In order that the statements in the *Q* sorts have precise and uniform meaning for all evaluators, definitions of terms that I thought might still prove ambiguous were read to the psychologists and the psychiatrists. During the course of the psychiatrists' second and third evaluations, a number of instances arose in which a psychiatrist commented about the possible ambiguity of some term that had not been defined. I could not then decide upon a definition for that term since it might not have coincided with the way the psychiatrist had implicitly defined the term to himself during the one or two prior evaluations. Instead, the psychiatrist was asked what meaning he had ascribed to the term during his earlier appraisals and he was then told to continue using the term in the same way. The psychiatrist's definition was written down and then included among the definitions read to the six psychologists evaluating that particular case.

Control 1

In order to determine if the amount of agreement between the psychologists and the criterion was greater than chance, some estimate of chance agreement first had to be determined. It would have been incorrect to assume that chance agreement would lead to zero correlations, for it is possible that some of the items in the *Q* sorts apply to the great majority of young adult males while others have little or no applicability. In order to allow for this possibility, the following procedure was employed. Each psychiatrist was asked to do *Q*-sort evaluations for each area on some other young adult male patient whom he was seeing or whom he had seen in therapy in his private practice, and about whom he felt at least as knowledgeable as he was of the clinic patient at the time of the third evaluation. This made it possible to correlate the *Q*-sort evaluations for each of the six psychologists who evaluated the projective test data for a particular clinic patient with the psychiatrist's evaluations of the matched control case. These correlations then gave a measure of chance prediction.

Control 2

The fact that the clinic patients were all court cases—information that the psychologists were aware of before making their evaluations—while the control patients were not court cases, could have led to spuriously high correlations. In order to determine if the psychologists were aided by a significant degree of stereotype accuracy, an additional control had to be instituted. A sample of 10 of the 30 psychologists were asked to evaluate a third case. In order that the various subgroups be represented, the 10 psychologists included at least two from each of the experience groups and from each of the personal analysis groups. In evaluating the third case, the psychologists were misinformed by being told that the person to be appraised was a private therapy patient. All 10 of the psychologists later reported that they accepted the instructions given them and evaluated the patient with the set I wanted them to have.

In order to eliminate the effect of practice, three of the 10 psychologists did a case under control conditions before doing either of the cases under experimental conditions. Three others worked under control conditions between the first and second case done under experimental conditions. And four psychologists evaluated the control case after both evaluations under experimental conditions had been completed.

A second kind of balancing was also necessary. For each psychologist, the case done by him under control conditions was the same case as was done by one of the other psychologists among the ten, under experimental conditions. This allowed for a comparison to be made of the same cases evaluated under control and experimental conditions.

Reliability of the Criterion Measure

In order to measure the reliability of the psychiatrists' evaluations, the following procedure was instituted. Seven of the 10 psychiatrists, who were able to devote additional time, were asked to evaluate some private practice patient whom they had stopped seeing at least six months prior but of whom they had a good recollection. Then after a period of approximately two months, during which time they had no further contact with this patient, they were asked to again evaluate him. The two sets of evaluations were then correlated and a measure of reliability was arrived at. The psychiatrists were asked to select a patient they had not seen for at least six months since, with a more recent patient, there might have been a significant amount of forgetting between the first and second evaluations. But once six or more months had elapsed, it seemed unlikely that the further interval of two months would lead to a significant amount of additional forgetting. For this reason, the psychiatrists could not use their clinic patients for the reliability check since they had very recently terminated treatment with them.

RESULTS

In order to correlate the quantified assessments of psychologists and psychiatrists, the Method-of-Difference correlational formula, which is based on the equality of the variances and means of the two distributions being correlated, was used. The latter conditions are imposed by the forced Q distribution.

For every psychologist, his evaluations for each of the two cases that he did were correlated with the psychiatrist's four sets of evaluations (first, second, third, and corrected third evaluation). This procedure was followed for each of the six areas appraised. Thus a total of forty-eight correlations were computed for each of the thirty psychologists, making 1440 correlations in all. Each of these correlations was then transformed into a z equivalent. For each area every psychologist's two z 's (one for each case) were averaged and the average was treated as a score. Thus, for each of the 30 psychologists there were four mean scores, representing the degree of agreement with the psychiatrist's first, second, third, and corrected third evaluations, for each of the six areas. In order to arrive at a measure of the psychologist's over-all agreement with the criterion the mean scores for the six areas were averaged. Thus, there were four over-all mean scores for each of the thirty psychologists representing the degree of agreement with the psychiatrists at the time of the latter's four appraisals.

In order to test Hypothesis I, the mean scores representing the degree of agreement with the psychiatrists' corrected third evaluations were used since these were judged to represent their most considered and therefore most accurate judgments. The over-all mean scores and the mean scores for each of the six areas for each of the thirty psychologists were averaged. The results of these computations are the seven master mean scores presented in Table 1, Column 2. Column 1 of the same table contains the corresponding r values which represent in correlational terms, the average degree of agreement between the psychologists and the psychiatrists.

In order to establish a base of chance prediction with which these experimental mean scores could be compared, Control 1 was utilized. For each of the 30 psychologists, his evaluations for each of the two cases appraised were correlated with the psychiatrists' evaluations of the matched control cases. These correlation coefficients were also transformed into z equivalents. Every psychologist's two z 's (one for each case) were averaged for each of the six areas. An over-all score was computed in the same way as was described above. Then, the over-all mean scores and the mean scores for each of the six areas of each of the 30 psychologists were obtained. The result was a set of master control means which are presented in Table 1, Column 3. These represent a measure of chance agreement between the psychologists and the criteria.

Hypothesis I posited that the evaluations of the psychologists would agree with the evaluations of the psychiatrists to a degree significantly greater than chance. This hypothesis was tested by means of t tests in which the means reported in Columns 2 and 3 of Table 1 were compared. The differences between the two sets of scores are reported in Column 4 and the t values in Column 5. The over-all difference is significant, as are also the differences for Areas II, III, IV, V, and VI. For Area I, the difference is not significant. However, it should be noted that if the significance criterion was relaxed to the .10 level, this difference too would be significant.⁶

⁶ In judging p values in this study, the usually accepted criterion of .05 is being taken as indicating significance, and it is only when this level is reached that a hypothesis can be considered as having received solid support. However, in a number of its aspects this study is investigating problems that have not been dealt with in prior published research. These problems include the degree to which psychologists can make accurate evaluations from projective test data for a number of delineated areas of functioning involving different levels of personality, differences in predictive ability of psychologists for these areas in terms of experience and personal analysis variables, and the ability of psychologists to offer psychotherapists information about patients for these areas that is not apparent from early therapy sessions. Since

Before these results can be taken as supporting Hypothesis I, the results of Control 2 must be evaluated. In order to determine if there were significant differences between the psychologists' correlations and the psychiatrists', when the former group knew that the patients were court cases compared to when they were not aware of this fact, t tests were performed. The t values were all extremely small and none were significant.⁷ Thus, it can be assumed that the significant positive correlations between the evaluations of the psychologists and the psychiatrists, reported in Table 1, were not dependent on the psychologists' knowledge that the patients who were appraised were court referrals. Hypothesis I can therefore be considered as supported for the over-all evaluations of the psychologists and for their evaluations in the areas of Motivating Needs and Affects, Character Traits, Diagnosis and Symptoms, Interpersonal Behavior, and Infancy and Childhood Perceptions. In the area of Defenses, there is a tendency in the hypothesized direction.

Hypothesis II posited that the more experienced the psychologists, the more accurately they would be able to evaluate projective test material. This was tested by a one-dimensional analysis of variance design as presented in Edwards (1950, p. 186), in which the three experience groups were compared. The mean scores for each experience group and the summaries of the analyses of variance are presented in Table 2. The F values did not reach one either

new horizons are being investigated, it would be particularly unfortunate if a Type II error was to be made. For this reason, mention shall also be made of results that would be significant if the significance criterion was relaxed to the .10 level. Such results should be considered as indicating a tendency in a particular direction and should be regarded with a good degree of tentativeness. They should be thought of as providing leads to be followed up in future investigations.

⁷ In this instance, and in all future instances where t values and F values are described as insignificant without further qualification, it should be understood that significance would not be reached even if the criterion was relaxed to the .10 level.

for the over-all differences or for the differences in any of the six areas, and thus there is no justification for abandoning any of the null hypotheses. Therefore, Hypothesis II received no support from the analysis of the data.

Hypothesis III posited that the psychoanalyzed psychologists would evaluate projective test material more accurately than the nonpsychoanalyzed psychologists. Again, the correlations of the psychologists' evaluations with the psychiatrists' corrected third evaluations were used. Table 3 presents the mean scores for the two groups, the differences between the two groups, and the *t* values obtained through tests of significance. The over-all difference between the two groups was in the predicted direction and significant. In all areas except Area V, the differences are again in the predicted direction, but they are not significant. However, if the significance criterion was relaxed to the .10 level, the differences for Areas II, IV, and VI would be significant. Therefore, Hypothesis III is supported

when the over-all evaluations of the two groups are considered. When the areas of Motivating Needs and Affects, Diagnosis and Symptoms, and Infancy and Childhood Perceptions are considered separately, there are tendencies in the hypothesized direction. For the areas of Defenses, Character Traits, and Interpersonal Behavior, no such tendencies are present.

Hypothesis IV posited increasing agreement between the psychiatrist's successive evaluations of his patient and the evaluations of the psychologists. Hypothesis V posited that this increasing agreement would be greater for the more experienced psychologists. In order to test both these hypotheses a comparison of validity correlations using the psychiatrists' first, second, and third evaluations was made for the psychologists as a total group and for psychologists divided into experience groups. Correlations were transformed into *z* equivalents and subjected to a "repeated measurements of the same subjects" analysis of variance design as outlined by Edwards

TABLE 1
MEAN SCORES FOR VALIDITY AND MEAN SCORES FOR CONTROL 1
AND *t* TESTS FOR DIFFERENCES BETWEEN THEM

Areas	Average Correlations for Experimental Cases Based on Z-Transformations	Mean Scores		Diff.	<i>t</i> Value
		Experimental Cases	Control Cases		
Defenses	.16	.1567	.0820	.0747	1.71 ^d
Motivating Needs and Affects	.23	.2260	.0740	.1520	3.91 ^a
Character Traits	.41	.4430	.1653	.2777	4.69 ^a
Diagnosis and Symptoms	.40	.4293	.3187	.1106	3.06 ^b
Interpersonal Behavior	.14	.1433	.0333	.1100	2.24 ^c
Infancy and Childhood Perceptions	.31	.3217	.1770	.1447	2.78 ^b
Over all	.28	.2867	.1417	.1450	5.47 ^a

^a Significant at .001 level.

^b Significant at .01 level.

^c Significant at .05 level.

^d Would be significant if criterion was relaxed to .10 level.

TABLE 2
MEAN SCORES FOR THE THREE EXPERIENCE GROUPS AND
SUMMARIES OF THE ANALYSES OF VARIANCE

Areas	Mean Scores			Source of Variance	df	MS	F value
	Group I	Group II	Group III				
Defenses	.189	.138	.143	Between Within	2 27	.0079 .0224	*
Motivating Needs and Affects	.211	.201	.266	Between Within	2 27	.0123 .0224	*
Character Traits	.500	.454	.375	Between Within	2 27	.0401 .0489	*
Diagnosis and Symptoms	.443	.447	.398	Between Within	2 27	.0074 .0204	*
Interpersonal Behavior	.151	.155	.124	Between Within	2 27	.0026 .0283	*
Infancy and Childhood Perceptions	.275	.354	.336	Between Within	2 27	.0172 .0286	*
Over all	.295	.292	.274	Between Within	2 27	.0013 .0102	*

* Error estimate larger than variance estimate for the effect.

TABLE 3
MEAN SCORES FOR THE TWO PERSONAL ANALYSIS GROUPS
AND *t* TESTS FOR DIFFERENCES BETWEEN THEM

Areas	Mean Scores		Diff.	<i>t</i> Value
	Psychoanalyzed Group	Nonpsychoanalyzed Group		
Defenses	.181	.137	.044	.59
Motivating Needs and Affects	.300	.195	.105	1.92 ^b
Character Traits	.502	.390	.112	1.50
Diagnosis and Symptoms	.484	.391	.093	2.03 ^b
Interpersonal Behavior	.111	.167	-.056	.75
Infancy and Childhood Perceptions	.392	.279	.113	1.72 ^c
Over all	.328	.260	.068	2.15 ^a

^a Significant at .05 level.

^b Would be significant if criterion was relaxed to .06 level.

^c Would be significant if criterion was relaxed to .10 level.

(1950, p. 284). The results of these analyses are presented in Table 4.

In regard to over-all differences and for Areas I, IV, and VI, the F values are not significant so that there is no justification for rejecting the null hypothesis. For Areas III and V, the F values for differences between the three sets of scores are significant so that the null hypotheses can be rejected. For Area II, the F value would be considered significant if the significance criterion was relaxed to the .10 level. For the last three areas mentioned, t tests were performed.

For Area II, the only significant t value was for the comparison of the first and the second set of scores. The mean for the

second set was the greater so that the results were in the anticipated direction. For Area III, significant t values were found when the first and third sets of scores were compared and when the first and second sets of scores were compared. In both instances, the mean of the first set was the lower of the two so that the results were again in the hypothesized direction. For Area V, a significant t value was obtained when the first and third sets of scores were compared. When the second and third sets of scores were compared, the t value would be considered significant if the significance criterion was relaxed to the .10 level. However, in both these instances the mean of the third evaluation was the lower of the two so that these

TABLE 4
SUMMARIES OF ANALYSES OF VARIANCE FOR DIFFERENCES IN AGREEMENT WITH PSYCHIATRISTS
AT THREE TIME INTERVALS AND INTERACTION EFFECT OF THESE DIFFERENCES
WITH DIFFERENCES FOR EXPERIENCE VARIABLE

Areas	Source of Variation	df	MS	F Value
Defenses	Between trials	2	.0049	1.56 ^d
	Interaction trials \times Experience groups	4	.0095	
	Pooled interaction (Error term)	54	.0061	
Motivating Needs and Affects	Between trials	2	.0217	2.52 ^e ^d
	Interaction trials \times Experience groups	4	.0036	
	Pooled interaction (Error term)	54	.0086	
Character Traits	Between trials	2	.0629	8.50 ^a ^d
	Interaction trials \times Experience groups	4	.0037	
	Pooled interaction (Error term)	54	.0074	
Diagnosis and Symptoms	Between trials	2	.0202	1.98 ^d
	Interaction trials \times Experience groups	4	.0029	
	Pooled interaction	54	.0102	
Interpersonal Behavior	Between trials	2	.0250	3.79 ^b ^d
	Interaction trials \times Experience groups	4	.0034	
	Pooled interaction (Error term)	54	.0066	
Infancy and Childhood Perceptions	Between trials	2	.0077	^d
	Interaction trials \times Experience groups	4	.0022	
	Pooled interaction	54	.0094	
Over all	Between time intervals	2	.0013	^d
	Interaction trials \times Experience groups	4	.0004	
	Pooled interaction (Error term)	54	.0020	

^a Significant at .01 level.

^b Significant at .05 level.

^c Would be significant if criterion was relaxed to .10 level.

^d Error estimate larger than variance estimate for the effect.

results are the opposite of what was hypothesized.

Thus, Hypothesis IV was supported in the area of Character Traits. In the area of Motivating Needs and Affects, there was a tendency in the hypothesized direction. In regard to over-all differences and differences in the areas of Defenses, Diagnosis and Symptoms, Infancy and Childhood Perceptions, and Interpersonal Behavior, no such tendencies were present. For the last mentioned area, there were results that were the opposite of those predicted. That is, the psychologists showed decreasing agreement with the psychiatrists' successive evaluations.

Hypothesis V receives no support from this study either for over-all differences between the three experience groups or for differences in each of the six areas treated separately. Since none of the interactions between experience groups and increases in scores at the three time intervals are signifi-

cant (Table 4), the more experienced psychologists were not increasingly more accurate than the less experienced.

Hypothesis VI posited that the psychoanalyzed psychologists would show a greater degree of increasing agreement with the psychiatrists' successive evaluations than the nonpsychoanalyzed. Again, a "repeated measurements of the same subjects" analysis of variance design was employed; this time the two variables were the increase in scores at the three time intervals and the two personal analysis groups. The relevant results of these analyses are presented in Table 5. The interaction term is the appropriate one for judging differences between the two groups. In regard to over-all differences, the *F* value is not significant. For Areas II, III, IV, and V, the *F* values for interaction are not significant so that there is no reason to reject the null hypothesis. In Area I, the *F* value is significant, indicating that the null hypothesis can be dis-

TABLE 5

SUMMARIES OF ANALYSES OF VARIANCE FOR INTERACTION EFFECT OF DIFFERENCES IN AGREEMENT WITH PSYCHIATRISTS AT THREE TIME INTERVALS AND DIFFERENCES FOR PERSONAL ANALYSIS VARIABLE

Areas	Source of Variation	<i>df</i>	<i>MS</i>	<i>F</i> Value
Defenses	Time intervals \times Personal analysis	2	.0192	3.37 ^a
	Pooled interaction (Error term)	54	.0057	
Motivating Needs and Affects	Time intervals \times Personal analysis	2	.0124	1.70
	Pooled interaction	54	.0073	
Character Traits	Time intervals \times Personal analysis	2	.0040	e
	Pooled interaction	54	.0070	
Diagnosis and Symptoms	Time intervals \times Personal analysis	2	.0015	e
	Pooled interaction	54	.0098	
Interpersonal Behavior	Time intervals \times Personal analysis	2	.0029	e
	Pooled interaction	54	.0069	
Infancy and Childhood Perceptions	Time intervals \times Personal analysis	2	.0208	2.42 ^b
	Pooled interaction	54	.0086	
Over all	Time intervals \times Personal analysis	2	.0028	1.47
	Pooled interaction	54	.0019	

^a Significant at .05 level.

^b Would be significant if criterion was relaxed to .10 level.

^c Error estimate larger than variance estimate for the effect.

carded. In Area VI, the F value would be significant if the significance criterion was relaxed to the .10 level. Thus, t tests were performed for these two areas so that the results could be further analyzed.

For Area I, the psychoanalyzed psychologists showed the hypothesized increasing agreement between the second and third sets of scores to a significant degree. Between the first and third sets of scores they would have shown the hypothesized significant increasing agreement if the significance criterion was relaxed to the .10 level. The nonpsychoanalyzed psychologists showed no such increase. For Area VI, the psychoanalyzed psychologists showed significant increased agreement between the first and second sets of scores, while the nonpsychoanalyzed group did not.

Hypothesis VI, then, receives support for the area of Defenses while in the area of Infancy and Childhood Perceptions there is a tendency in the hypothesized direction. In terms of the over-all differences between the two personal analysis groups and for the areas of Character Traits, Diagnosis and Symptoms, Interpersonal Behavior, and Motivating Needs and Affects, no such tendencies are present.

In addition to the statistics already presented which bear upon the six hypotheses, additional statistical work was undertaken. One problem that was investigated was related to the original grouping of psychologists along the personal-analysis dimension. The nonpsychoanalyzed group consisted both of psychologists who had received some form of treatment, other than Freudian psychoanalysis (Subgroup B₁), and of those who had received no form of treatment (Subgroup B₂). Comparisons of these two subgroups were made regarding the amount of agreement between their evaluations and the corrected third evaluations of the psychiatrists to determine if, in fact, they performed in the similar way that the original grouping implicitly suggested that they would. The mean scores for both groups and the t values for the differences between the means were derived. None of the t values were significant, indicating that there is no reason to reject the null hypoth-

esis regarding the differences between the two groups. Thus, the decision to consider these two subgroups together in comparing them with the psychoanalyzed psychologists is supported empirically.

In order that three-way comparisons of the psychoanalyzed group (Group A), Subgroup B₁, and Subgroup B₂ could also be made, the mean scores for each of these three groups were computed. The scores for Group A were higher than the scores for Subgroup B₁ and Subgroup B₂ for over-all differences and for Areas II, III, IV, and VI. However, when the differences were evaluated statistically, because of the breakdown of the psychologists into three rather than two groups, the differences between the psychoanalyzed and nonpsychoanalyzed psychologists were less apparent. When a two-way comparison was made, the over-all difference was significant; when a three-way comparison was made, it was not. However, it should be noted that if the significance criterion was relaxed to the .10 level, this difference, too, would be significant. Although when a two-way comparison was made it was stated that this relaxation of the significance criterion would result in significant differences for Areas II, IV, and VI, in the three-way comparison this would only be true for Area II.

Other statistics of interest center about interpsychologist reliability. In order to determine the degree of agreement between psychologists evaluating the same case, the same formula was employed as was utilized in computing validity coefficients. Each psychologist's set of evaluations (one for each of the six personality areas) were correlated with the other five psychologists appraising the same case. This yielded 15 sets of correlations per case or a total of 150 sets of correlations for all 10 cases. The correlation coefficients were transformed into z equivalents, means computed and these mean z 's transformed back into correlation coefficients. The coefficients were .27 for Defenses, .25 for Motivating Needs and Affects, .44 for Character Traits, .44 for Diagnosis and Symptoms, .21 for Interpersonal Behavior, .38 for Infancy and

Childhood Perceptions, and .34 over all. These coefficients represent the average degree of agreement among psychologists who evaluated the same projective test material.

Just as the psychologists' validity coefficients could be compared for the different experience groups and for the two personal analysis groups, so could their reliability coefficients also be compared. The first problem to be investigated in this regard was whether degree of experience played a part in the amount of interpsychologist agreement. Did the more experienced psychologists agree more among themselves than the less experienced psychologists? The correlations of the two psychologists from Group I who did the same case, those from Group II, and those from Group III were pertinent in this regard. Since there were 10 cases, 10 sets of correlations for each of the three experience groups were compared. These comparisons were made by means of a one-dimensional analysis of variance design. None of the *F* values either for over-all differences or for any of the six areas considered separately were significant. Thus, there is no evidence from this research to suggest that the degree of experience of a psychologist plays a role in determining how much his evaluation of a case will agree with another psychologist's evaluation who is approximately at the same level of experience.

A second and even more crucial question was asked regarding interpsychologist reliability in relation to the experience dimension. Do the more experienced psychologists agree more among themselves than they agree with the less experienced psychologists? In order to answer this question, only the evaluations of the 20 psychologists from the two extreme experience groups were considered. For each of the 10 cases, the correlations, representing agreement of each of the psychologists in Group I with both of the psychologists in Group III were utilized. For all 10 cases, this amounted to 40 sets of correlations. Mean scores were computed and compared with the mean scores representing agreement among the most experienced psychologists with each other. These comparisons were made in *t*

tests, and none of the differences were significant. Thus, this study offers no support for the idea that highly experienced psychologists will agree more among themselves than they will agree with less experienced psychologists in projective test evaluations.

To study interpsychologist reliability in relation to the personal analysis dimension, comparisons were made in the following way. The evaluations of every psychoanalyzed psychologist were correlated with the evaluations of every other psychoanalyzed psychologist who did the same case. Similarly, the evaluations of every nonpsychoanalyzed psychologist were correlated with the evaluations of every other nonpsychoanalyzed psychologist who did the same case. This produced a total of 12 sets of interpsychologist correlations among the psychoanalyzed group and 35 sets of interpsychologist correlations among the nonpsychoanalyzed group. These correlations were transformed to *z* equivalents, means were computed for each of the two groups, and differences were tested by *t* tests. The over-all difference between the two groups was significant, as were also the differences for the areas of Defenses, Character Traits, and Infancy and Childhood Perceptions. In the area of Motivating Needs and Affects, the difference would have been significant if the significance criterion was relaxed to the .10 level. In all four of these individual areas, as for over-all differences, the mean scores for the psychoanalyzed psychologists were the higher. In the areas of Diagnosis and Symptoms and Interpersonal Behavior, there were no significant differences between the two groups.

Another type of reliability score that can be reported is that of each of the seven psychiatrists who evaluated a private patient at two intervals, not having seen the patient during the interim. The two evaluations of each psychiatrist were correlated with each other and the coefficients were transformed into *z* equivalents. Means were then computed for the *N* of seven, and these in turn were transformed back into correlation coefficients. The coefficients were .70 for Defenses, .64 for Motivating Needs and Affects, .76 for Character Traits, .79 for

Diagnosis and Symptoms, .70 for Interpersonal Behavior, .76 for Infancy and Childhood Perceptions, and .73 over all.

The final data to be reported relate to individual differences in validity coefficients for the six psychologists who evaluated the same case. The correlations of each psychologist with the criterion were utilized for this purpose. The spreads of these correlations for each case were determined, the term "spread" referring to the difference between the highest and lowest of the six correlations for a particular case. Then the means of these spreads were computed. These means were .45 for Defenses, .45 for Motivating Needs and Affects, .42 for Character Traits, .42 for Diagnosis and Symptoms, .69 for Interpersonal Behavior, and .44 for Infancy and Childhood Perceptions.

DISCUSSION

An extensive and detailed discussion of the findings of this study can be found elsewhere (Silverman, 1958). Only the most important issues will be dealt with here and in some instances in summary form.

The findings demonstrating that the group of 30 psychologists was able to appraise projective test material with a degree of accuracy significantly greater than would be expected by chance, while not sufficient for great rejoicing on the part of clinicians, do have some importance. For the only personality areas in which projective tests, used holistically, have been adequately demonstrated to have even this much validity in the past, have been those of Diagnosis and, to a lesser degree, Character Traits, as these areas have been defined in the current investigation. The results of the current research support the findings for these two areas. In addition, the current research has demonstrated that clinicians using projective techniques can evaluate patients to a degree greater than chance in three additional areas. In the areas of Motivating Needs and Affects and Infancy and Childhood Perceptions, these results are particularly noteworthy. For, as far as I could discern from published research, this is the first time that it has been demonstrated in a controlled, quantified, and objectively evaluated study that clinical psychologists utilizing projective techniques can make inferences about the underlying motivating forces in patients that are congruent with the inferences made by psychiatrists utilizing data revealed in psychotherapy. The positive results in the areas of Interpersonal Behavior and Infancy and Childhood Perceptions are noteworthy in another way since, for these two areas, many psychologists, including

a number participating in this study, stated that they did not think accurate appraisals based on projective tests could be made.

By usual standards the mean validity coefficients reported in Table 1 for the six areas would be classified as ranging from "low" to "moderate." However, such standards would have little meaning for this study since they are based on the assumption that chance correlation is .00 and maximum correlation is 1.00. In this study, neither of these assumptions is justified. The correlations attained by Control 1 (Table 1) represent chance agreement and all of these are above .00. These "true" bases should be kept in mind in judging the size of the mean validity correlations, though it can be noted that for five of the areas, the coefficients are low.

In judging the size of the validity coefficients, a second consideration is the actual maximum above which no psychologist's correlations could be expected to go. Just as .00 could not be taken as chance correlation, 1.00 cannot be taken as the true maximum for this would imply that the criterion measure was completely valid. This, of course, no one, least of all the psychiatrists themselves, would claim. Just what the "true" maximum might be is extremely difficult to estimate, for by what could we judge the accuracy of the psychiatric evaluations? However, one aspect of the validity of the psychiatrists' evaluations that was measured was the reliability of their judgments. While reliability is just one small aspect of validity, it does set a limit beyond which validity cannot go. The intrapsychiatrist reliability coefficients reported earlier can be used to estimate indices of reliability. These indices (which are the square roots of the intrapsychiatrist reliability coefficients), can be considered as approximating the upper limits beyond which the psychologists' evaluations for a particular area could not go. The mean indices were .84 for Defenses, .80 for Motivating Needs and Affects, .87 for Character Traits, .89 for Diagnosis and Symptoms, .84 for Interpersonal Behavior, .87 for Infancy and Childhood Perceptions, and .85 over all.

In judging the validity coefficients, the reader should also keep in mind the following artifacts that may have lowered them so that they represent an underestimate of the accuracy with which psychologists are actually able to evaluate projective test material:

1. Unlike the usual clinical procedure, the psychologists did not test the patient whom they were to evaluate, but had to depend on the notes of another person.

2. Failure to include an intelligence test such as the Wechsler-Bellevue in the test battery also made the psychologists' task different from what it usually is in clinical practice and consequently may have prevented them from operating with their usual effectiveness. The oversight of not including a highly structured test such as this may have had particularly costly consequences in the current

study since two of the areas to be evaluated dealt with the subject's actual manner of coping with the world (Character Traits and Interpersonal Behavior).

3. In validation research, the participating psychologists, by necessity, must interpret the test material independently of other sources of data. In clinical practice, on the other hand, such interpretations can be made in the context of information gathered from interviews or case history abstract. The advantage of having such data available, carrying with it the implication that more justice could be done to the projective test material itself, has been elaborated upon by Schafer (1954).

4. The use of *Q* sorts for evaluation handicapped the psychologists in a number of ways. For one thing, this was a new and different way of recording psychological evaluations for many of the psychologists and their unfamiliarity with the procedure may, in itself, have impaired the effectiveness with which they operated. Secondly, unlike the psychological report, the *Q*-sort procedure compelled the evaluators to appraise a great many aspects of the patient in question, regardless of whether or not they felt the test material yielded pertinent data. Moreover, while in a psychological report interpretations can be made with varying degrees of certainty by qualifying some statements as high level inferences and others as speculations, there is no place for this with *Q* sorts. Items are only judged in terms of the importance they have in the patient's personality and not in terms of how confident the evaluator feels in judging the particular item. And finally, just as the forced-choice aspect of *Q* sorts may have compelled the psychologist to go out on a limb, they conversely limit him in those aspects of the patient that he can appraise.

5. The fallibility of the psychiatrists must have contributed in some measure to the fact that the correlations were not higher than they were. One way of reducing errors caused by such variables as countertransference distortions might be to use, for the criterion measure, the pooled evaluations of a group of psychotherapists, each of whom is presented an account, as close to verbatim as possible, of the actual treatment sessions as well as additional data available such as results from social service investigations. A procedure similar to this was actually employed in the studies of Fisher (1952) and Little and Shneidman (1955) and the increase in validity coefficients was marked. The adequacy of the criterion measure might also be increased by enlisting the cooperation of more highly experienced and skilled psychotherapists, such as training analysts, by utilizing as *Ss* patients who are being seen in private treatment and thus would probably be better motivated to produce material, and by utilizing patients who are undergoing a more intensive and extensive type of psychotherapy, such as psychoanalysis.

6. Semantic confusion seemed to have caused a large segment of apparent disagreement between

psychologists and the criterion. My attempts to deal with this problem, which were described earlier, were far from fully effective for two reasons. First, a number of words which neither I nor the psychiatrists believed to be ambiguous enough to warrant definitions were shown later to have been implicitly defined in different ways by different evaluators. This came to light in discussions with the evaluators after the evaluations had been completed. It was of particular interest to note that some of the terms implicitly defined differently were nontechnical ("optimistic" and "grandiose" for example) and find their way into the parlance of people other than psychologists and psychiatrists. Thus it appears that this semantic problem extends beyond the use of psychiatric terminology, and to secure uniformity in understanding, virtually every key word should be defined in a study such as this.

Secondly, in some instances, the evaluator did not really accept the definition that he was given. He would acknowledge the definition by nodding his head, but from later discussion it was evident that he had lapsed back into thinking of the term as he had understood it in the past while doing the sorting.

7. Related to this problem of semantic confusion was the difficulty caused by the broadness of many of the concepts used. For example, interpersonal behavior, such as withdrawing from heterosexual relations, defying authority, or competing with peers, can apply in certain ways to a patient and not in others. Sometimes postevaluation informal discussion revealed that one evaluator selected certain aspects of a patient's behavior to base his evaluation on while ignoring other aspects. If the psychologist and the psychiatrist happened to focus on different aspects of the patient in responding to an item, pseudo-discrepancies between their two evaluations were apt to arise. There is no easy solution to this problem. All items could be made highly specific. Thus, instead of the statement "the patient withdraws from sexual contact with those of the opposite sex," a number of different items could be specified, one related to sexual contact with older females, another to such contact with younger females, and so on. While this type of specificity would greatly reduce the number of ways in which the statements could be viewed, it is questionable if such highly specific statements can be made with confidence from material culled from psychological tests.

8. Another probable cause of spurious differences between the projective evaluators and the psychiatrists was differences in orientation. There was much variation within both the psychologist and the psychiatrist groups. In the former, 14 described their orientation as predominantly Freudian, 6 as predominantly that of the Interpersonal School, and 10 stated their position was eclectic, with both Freudian and "interpersonal" concepts playing a major role in their theoretical thinking. The psychiatrists included two who were trained

in Freudian psychoanalysis, four who were trained at the William Alanson White Institute, and four at the Columbia Psychoanalytic Clinic for Training and Research. When a psychiatrist of one orientation evaluated the same case as a psychologist with a different orientation, differences based on their orientations were apt to occur in the areas of Motivating Needs and Affects and Infancy and Childhood Perceptions. Case discussions that I have had with various members of the Psychiatric Clinic staff have led me to believe that it is in these two areas that differences between schools come into sharpest focus.

Note should also be taken of the size of the reliability coefficients representing the average degree of agreement among psychologists who evaluated the same projective test material. In general, these correlations are disappointingly small and much lower than one expects reliability coefficients to be. However, it would not be possible to say how much this was due to artifacts rather than real disagreement among the participating psychologists. Many of the artifacts just discussed as lowering the validity coefficients may have lowered the reliability coefficients as well.

In light of the low interpsychologist reliability, it is hardly surprising that much variability was found in comparing the validity coefficients of the six psychologists who evaluated the same case. The large spreads of correlations reported earlier imply that the differences in ability, with which the various participants were able to operate, were great, particularly when considered in the context of the upper and lower limits, above and below which the correlations would not be expected to extend (control correlations and indices of reliability, respectively). These findings are in keeping with the results of the other two studies using *Q* sorts (Fisher, 1952; Little & Shneidman, 1955) and three other studies that utilized other methods of appraisal (Chamber & Hamlin, 1957; Samuels, 1952; Symonds, 1955), all of which reported great individual variation in the effectiveness with which different psychologists were able to make evaluations. It seems justified to conclude from these various results that a projective test evaluation is far from a mechanical act which proceeds along the lines of an X-ray reading in which any number of technicians, with a certain minimum amount of training, will arrive at the same results. Instead, drawing inferences from projective techniques would appear to be largely an art which depends on the skill of the individual clinician. Thus, the question often asked "How great is the validity of projective technique appraisals?" can only be answered if it is first specified *who* is making the appraisal.

In regard to the experience variable, there were no differences among the three experience groups that were significant, and many of the small differences that were obtained went in the unpredicted direction. The various artifacts that have already been alluded to as probable causes of spuriously

low validity coefficients would not be pertinent in this regard since there is no reason to suspect that they operated systematically, affecting one experience group more than the others. One possible exception to this might be in the shortcomings of the criterion measure. It could be argued that the appraisals of the psychiatrists were so inadequate that their evaluations included only information about the patients that was relatively obvious and that required relatively little skill in discerning from the projective test material. The more subtle aspects of the patients, which only the highly experienced psychologists could discern, according to this argument, were not reflected in the criterion evaluation. However, this argument loses its cogency when the interpsychologist reliability coefficients are noted. It has already been noted that the most experienced psychologists did not agree any more among themselves than did the least experienced psychologists. Moreover, the most experienced psychologists agreed as much with the least experienced psychologists as they agreed with each other. The absence of differences here cannot be blamed on any deficiencies in the criterion measure.

A second way in which these findings could be challenged is in terms of the evaluative task itself. It could be argued that the kind of evaluation required was so crude that differences in the effectiveness with which different psychologists operated could not be high-lighted. However, this argument, too, would be refuted in light of two findings. One is that the range of correlations for those clinicians evaluating the same case has, for all areas, been judged to be very large. Secondly, significant differences between groups for both validity and interpsychologist reliability coefficients have been found to exist along the personal analysis dimension. Thus, the evaluative task in the current study is one in which differences between psychologists can come to light, if such differences exist.

When all these findings are considered, the lack of any evidence even suggesting that experience level plays a role in determining the effectiveness with which clinical psychologists evaluate projective material, seems to me to be of major import. For as far as I could discern from published research, this is the first projective technique validation study that has systematically investigated differences along the experience dimension where the psychologists were required to appraise aspects of personality that they appraise in their everyday clinical practice. The failure of other researchers to investigate this important question may well be due to their taking for granted the idea that degree of experience is of crucial importance. That this was the case with at least some investigators is suggested by the fact that, in certain studies (Fisher, 1952; Little & Shneidman, 1955), highly experienced psychologists were deliberately chosen as the projective technique evaluators as if there was the underlying assumption that in this way,

positive results were more apt to be obtained. While such an assumption is consistent with what logic would lead us to expect, the findings of this study tell an entirely different story.

The negative findings for differences along the experience dimension reported above are subject to two major qualifications. First, it is most important to bear in mind that the term "effectiveness" in this study refers only to the accuracy with which the clinical psychologist can make interpretations based on the projective data. This is just one way in which the competence of the projective tester is judged in actual clinical practice. Other skills such as organizing ideas, integrating various personality facets with each other, refining interpretations (that is, shading general concepts to fit a particular case), expressing oneself fluently, clearly, and with style, are also required, and these skills differentiate the usual report writing task from the Q-sorting task. These may be shown by future research to depend heavily on the experience level of the psychologist.

The fact that the participants in Group I were not only the most experienced psychologists, but were also persons with reputations for high competence in the field of projective testing, yet as a group, did not perform significantly more effectively than Groups II and III, also becomes understandable in light of the qualification as to what the term "effectiveness" applies to in the study. For the reputations of these individuals were earned through their teaching, writing, and supervisory work. The skills that underlie these functions, as well as the report writing skills mentioned above, can all be present to a much greater degree in these psychologists than in the others, yet not the accuracy with which they make interpretations.

The second qualification to the findings reported above is that their applicability is limited to the various experience levels characterizing the psychologists who participated in this study. This qualification could be of crucial importance since the least experienced of the three groups could hardly be said to be composed of complete novices. This group included psychologists who had between one and three years of clinical experience, where this experience was preceded by or accompanied by an average of 4.7 lecture, practicum, and supervisory courses in projective testing. Thus, most of the members of Group III had a moderate amount of contact with projective techniques. It may well be that the kind of ability required in the evaluative task in this study required, as a minimum for its effective execution, just such a moderate amount of contact with projective techniques and that beyond this minimum, further experience was of no assistance, with the relative degree of effectiveness with which different psychologists operated, depending on other factors. If this is true, then we would expect real beginners—that is, individuals who have just completed one or two courses in projective testing or who have started on, or even just completed an internship—

to operate significantly less effectively than the psychologists in Group III who participated in this study. This hypothesis would be simple enough to test in future research using the same Q sorts and the same criterion measures that were employed here.

The results reported in Table 3 indicated that the hypothesis positing greater validity for the evaluations of the psychoanalyzed psychologists was supported to some extent. The possibility that the orientation of the evaluator could have influenced his ratings in two areas (Motivating Needs and Affects, and Infancy and Childhood Perceptions), has already been discussed. Could the higher over-all validity coefficients of the psychoanalyzed psychologists, therefore, reflect, in part, the greater similarity between their orientations and the orientation of the psychiatrists? This question can be answered in the negative. The orientation of all the psychoanalyzed psychologists was Freudian. This was only true of four non-psychoanalyzed psychologists. Six others described their orientation as that of the interpersonal school. The other ten described their orientation as being eclectic, drawing heavily from both these frames of reference. Since the orientation of eight of the ten psychiatrists was non-Freudian, with at least four and probably all eight influenced by the thinking of the interpersonal school, to the extent that orientation played a part in determining results, the psychoanalyzed psychologists were at a disadvantage. Thus the fact that the over-all difference between the two groups was significant, suggests that this difference represents an underestimate of the greater effectiveness of the psychoanalyzed psychologists.

The differences between the two groups came into even sharper focus when certain aspects of interpsychologist reliability were considered. These involved a comparison of the degree of agreement among the psychoanalyzed psychologists doing the same case with the degree of agreement among the nonpsychoanalyzed psychologists doing the same case. For over-all scores, the difference between the mean reliability coefficients of the two groups was .13 compared to a difference of .07 for their mean validity coefficients. The former difference is significant at the 1% level, while the latter is significant at the 5% level. For individual areas, the differences for reliability were also much greater than the differences between the validity coefficients in the following three areas:

Defenses. Differences in validity coefficients .04 (insignificant)—differences in reliability coefficients .14 (significant at the 5% level).

Character Traits. Difference in validity coefficients .11 (insignificant)—difference in reliability coefficients .29 (significant at the 1% level).

Infancy and Childhood Perceptions. Difference in validity coefficients .11 (significant only if the significance criterion is relaxed to the .10 level)—difference in reliability coefficients .16 (significant at the 5% level).

Two explanations will be posited which I believe are most apt to account for the fact that the differences between the psychoanalyzed and non-psychoanalyzed groups were greater for reliability than for validity. One of these revolves around the inadequacies of the criterion measure. It may be that these inadequacies minimized the actual degree of difference between the two groups as measured by their validity coefficients. The differences in reliability coefficients, on the other hand, not being dependent on the criterion measure, would have then reflected more accurately just how much more effective was the psychoanalyzed group.

The second explanation would revolve around the relationship between orientation and the personal analysis dimension. Since orientation could have affected ratings in some areas, two psychologists evaluating the same case who were of the same orientation could be expected to agree more with each other than two psychologists of different orientations. Inasmuch as all the psychoanalyzed psychologists were of one orientation while the nonpsychoanalyzed psychologists were heterogeneous in this regard, this is what might underlie much of the greater reliability in the former group.

In the area of Infancy and Childhood Perceptions, the latter explanation could be at least partially applicable. For, as was elaborated upon earlier, the degree of emphasis given to certain items in this area could well be a consequence of theoretical orientation. To the extent that this was the case, the higher reliability figures would reflect an artifact and not a more accurate index of difference between the psychoanalyzed and nonpsychoanalyzed groups.

In the area of Defenses, however, the explanation in terms of an artifact would have much less weight for there is no reason to believe that a particular orientation would lead a psychologist to weight one type of defense more heavily than another. There are differences between the two psychiatric schools of thought in their understanding of what the various defenses owe their origin to, but members of both groups would agree that these various defense mechanisms do exist. It is true that most of these concepts and the terms in which the concepts are expressed originated in Freudian psychoanalytic theory, so that psychologists with a Freudian orientation might have felt particularly comfortable dealing with them. However, these concepts and terms play a considerable part in the thinking of members of the interpersonal school as well. Thus, the fact that the difference in validity coefficients between the two groups was quite a bit lower than the difference in reliability coefficients, more likely reflects the shortcomings of the criterion measure. The latter difference, therefore, would more accurately reflect the degree of greater effectiveness of the psychoanalyzed group.

In the area of Character Traits, there is even less reason to believe that theoretical orientation of the psychologists influenced their evaluations. None of the traits mentioned has priority over others in the theory of either school. Moreover, the actual terms utilized were employed in the past long before the advent of psychoanalysis. Thus, for this area, the degree of difference between reliability coefficients is probably a more accurate estimate of the greater effectiveness of the psychoanalyzed group.

When differences between the psychoanalyzed and nonpsychoanalyzed groups for both validity and interpsychologist reliability are considered, the former group's effectiveness can be taken as notably greater than the latter's. To what is this greater effectiveness due? It has been my assumption that the answer to this question lies in certain benefits that accrue from undergoing psychoanalysis. It is possible, however, that this assumption is unjustified. Could it not be that some other factor led both to these psychologists' being more proficient in their work with projective techniques and to their entering this type of treatment? Would they then not have attained just as high correlations if they had not entered psychoanalysis? I believe that the first explanation which posited a cause and effect relationship is the more likely because of two kinds of evidence. The first is logical and has already been presented as the rationale underlying Hypothesis III, in which were given the reasons why undergoing such a form of treatment should increase the effectiveness with which psychologists interpret test data.⁸ The second is subjective, consisting of comments made by some of the psychoanalyzed psychologists to the effect that they believed that their effectiveness with projective tests increased as a result of psychoanalysis. However, neither of these reasons can be accepted as proof, for it is not unusual for logic to bow to empiricism, and moreover, it is possible that just as logical a rationale could be presented to support the alternative explanation. As for the psychologists' comments, such self-evaluations cannot always be trusted. It would be crucial in deciding this issue to compare the effectiveness of a group of psychologists just beginning psychoanalysis with the psychoanalyzed group in this study, using the same *Q* sorts and the same criterion measure. Only three of the twenty nonpsychoanalyzed psychologists had just begun psychoanalysis so that no such comparisons could be

⁸In addition to the explanation given in the Hypotheses section, the general therapeutic gains of psychoanalysis should be mentioned. To the extent that a psychoanalysis succeeds, anxiety and inhibitions should be minimized and more conflict-free energy should be at the analysand's disposal. These changes would lead to more effective functioning in the work area as well as others.

made here. Until such a step is taken, my explanation for the results rests on somewhat shaky ground.

Can only psychoanalyzed psychologists evaluate projective test material with a relatively high degree of effectiveness? An examination of the distribution of both validity and interpsychologist reliability scores among both the psychoanalyzed and nonpsychoanalyzed groups indicates that this is not the case. In regard to validity coefficients for over-all agreement, 37% of the nonpsychoanalyzed group achieved coefficients that were at least as high as the average coefficient for the psychoanalyzed group. Similarly, for over-all interpsychologist reliability, the coefficients for 33% of the nonpsychoanalyzed pairs were at least as high as the average reliability coefficient for the psychoanalyzed pairs. Thus, one can assume that some psychologists, either without treatment or as the result of successful psychotherapy, are in sufficient touch with intrapsychic forces that are usually unconscious and are sufficiently free of debilitating anxiety and conflict so that they function with a relatively high degree of effectiveness in evaluating projective test material.

Regarding the converse, there is the question of whether any psychoanalyzed psychologists were, in a relative sense, ineffective in their evaluations. The answer would be in the negative. In terms of over-all agreement with the criterion, the lowest validity coefficient of any psychoanalyzed psychologist was .25, which is only slightly lower than the mean validity coefficient for all 30 psychologists of .28. Similarly, the lowest reliability coefficient for interpsychologist agreement among the psychoanalyzed group is .29, not much lower than the .33 which represents mean reliability for interpsychologist agreement when all pairs of psychologists in the same personal analysis group are considered. Thus, if one can generalize from the present sample, it would seem that the fact that a psychologist has been psychoanalyzed offers some assurance that his projective test evaluations will be, in a relative sense, at least reasonably accurate.

The one finding that was diametrically opposite to what was hypothesized also deserves comment. In the area of Interpersonal Behavior, for the psychologists as a total group there was a decreasing degree of agreement between their evaluations and the psychiatrists' successive evaluations. I believe that the reason for this involved an artifact. It seems highly probable that the psychiatrists were not able to adhere to the instructions given them during their second and third evaluations, that they evaluate the patient in terms of how he was when he first entered treatment, excluding changes that took place during the period of therapy. Such instructions would be most difficult to follow. For there was no way in which a psychiatrist could be certain that the behavior the patient exhibited, or described after being treated for some time, was truly a new mode of interaction. The patient may simply have suppressed such action with the psy-

chiatrist and even suppressed describing the behavior until he felt more at home in the therapy situation. Overt behavior might well change, particularly during late adolescence and young adulthood. Thus, the psychiatrist may have rated items during the second and third evaluations, thinking that he was describing the patient as he originally was but in a truer light, while in reality he was describing new kinds of behavior.

Why, then, did not the same error occur in the other five areas that the psychiatrists evaluated? The answer would lie in the nature of the personality variables tapped in the other areas. Certainly Infancy and Childhood Perceptions of parental figures which, by definition, refers to the patient's past, could not change. For Diagnosis and Symptoms, no change would be anticipated since the nosological entities were stated in terms of vulnerability. Thus, even if a patient lost a particular symptom during such brief treatment, he would still be vulnerable to its recurrence in the future. Character Traits referred to lifelong reaction patterns, Motivating Needs and Affects to basic drives and fears, and Defenses to habitual responses to anxiety. Neither life events nor therapy of such short duration could be expected to bring about alterations in these three areas, even though behavioral changes might occur.

The failure to find the hypothesized differences along the experience dimension for increased agreement with the psychiatrists' successive evaluations, parallels the negative findings for comparisons of the three groups for their validity and interpsychologist reliability coefficients. Thus, here is another skill of the clinical psychologist, for which evidence is lacking, that degree of experience affects performance.

In terms of differences along the personal analysis dimension, for the area of Defenses the findings paralleled the results of comparisons between the two groups for validity and interpsychologist reliability coefficients, only in this instance, in a positive sense. Thus, another advantage that appears to accrue from a clinical psychologist undergoing psychoanalysis, subject to the reservation regarding cause and effect relationships made earlier, is that it increases his ability to detect central and hidden defensive patterns that are obscured from detection by psychotherapists until the patient has been in treatment for some time. Perhaps the psychoanalyzed psychologist's awareness of what is chaff and what is wheat in himself allows him to make the same differentiation in patients whose projective test material he evaluates.

In light of the consistent differences found between the psychoanalyzed and nonpsychoanalyzed psychologists in this study, future research on the differential ability with which clinical psychologists evaluate projective test data would do well to investigate this variable. The fact that the group designated as psychoanalyzed in this study included only those psychologists who had undergone Freudian psychoanalysis should be kept in mind,

however, since it is my belief that the combining of psychologists who receive various types of intensive or dynamic therapies into one group would obscure differences that actually exist.

SUMMARY

This study was undertaken to investigate the validity of projective technique evaluations. The patients whose test material was to be appraised were 10 young adult males who entered psychotherapy at the Psychiatric Clinic attached to the Court of Special Sessions of the City of New York. Each had administered to him a Rorschach, Thematic Apperception Test, House-Tree-Person Drawings, and the Most Unpleasant Concept Test before entering treatment.

The test material was evaluated by a group of 30 clinical psychologists. These psychologists were divided into three subgroups depending on degree of professional experience and two subgroups depending on whether or not they had undergone Freudian psychoanalysis. Each psychologist evaluated the protocols for two patients and each patient was evaluated by six psychologists.

The degree of validity of the psychologists' evaluations was determined by correlating them with evaluations made by the psychiatrists treating the patients after 35 or more therapy sessions. The psychiatrists were also asked to evaluate the patient at two earlier intervals during treatment to determine if there was increasing agreement between their evaluations and the psychologists' evaluations as the psychiatrists became better acquainted with the patients.

The method of evaluation for both psychologists and psychiatrists involved the use of *Q* sorts. There were six *Q* sorts, one for each of six personality areas to be appraised. These areas were: (a) Defenses; (b) Motivating Needs and Affects; (c) Character Traits; (d) Diagnosis and Symptoms; (e) Interpersonal Behavior; (f) Infancy and Childhood Perceptions of Parental Figures.

A summary of the results follows:

1. The psychologists as a total group were able to evaluate the projective test material

to a degree significantly greater than chance, both over all and for five of the six areas considered separately. For the sixth area (Defenses) there was a tendency in this direction.

2. There were no significant differences between the three experience subgroups in the size of their validity coefficients. There were also no significant differences in inter-psychologist reliability in terms of experience level.

3. Those psychologists who had undergone Freudian psychoanalysis had significantly higher over-all validity coefficients than those psychologists who had not received this form of treatment. While differences between these two subgroups were not significant for any of the six areas considered separately, in three of the areas there were tendencies in the same direction. There were also significant differences in interpsychologist reliability, both over all and for three of the areas considered separately. In all instances there was greater agreement among those psychologists who had undergone Freudian psychoanalysis than among those who had not.

4. The psychologists as a total group agreed to a significantly greater degree with the psychiatrists' later evaluations than with their earlier ones in the area of Character Traits. In the area of Motivating Needs and Affects there was a tendency in the same direction. In terms of over-all scores, and for the other four areas considered separately, there was no increasing agreement. In one of these areas (Interpersonal Behavior) the psychologists showed significantly decreasing agreement with the psychiatrists' successive evaluations.

5. In regard to the experience variable, there were no significant differences between the three subgroups in regard to their ability to show increasing agreement with the psychiatrists' successive evaluations.

6. In regard to the personal analysis variable, those psychologists who had undergone Freudian psychoanalysis showed significantly greater increasing agreement with the psychiatrists' successive evaluations in the area of Defenses than those psychol-

ogists who had not received this form of treatment. In the area of Infancy and Childhood Perceptions, there was a tendency in the same direction. For over-all scores, and for the other four areas consid-

ered separately, there were no differences between the two subgroups.

These findings were discussed and their implications for future research were commented upon.

REFERENCES

- BENJAMIN, J. D., & EBAUGH, F. G. The diagnostic validity of the Rorschach test. *Amer. J. Psychiat.*, 1938, **94**, 1163-1178.
- BIBRING, E. Psychoanalysis and the dynamic psychotherapies. *J. Amer. Psychoanal. Ass.*, 1954, **2**, 745-770.
- BUCK, J. N. The HTP test. *J. clin. Psychol.*, 1948, **4**, 151-159.
- CHAMBER, G. S., & HAMLIN, R. W. Validity of judgments based on "blind" Rorschachs. *J. consult. Psychol.*, 1957, **21**, 105-109.
- CRONBACH, L. J. Pattern tabulation: A statistical method for analysis of limited patterns of scoring with particular reference to the Rorschach test. *Educ. psychol. Measmt.*, 1949, **9**, 149-171.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
- FILER, R. N. The clinician's personality and his case reports. Unpublished doctoral dissertation, Univer. of Michigan, 1951.
- FISHER, LILLIAN. An investigation of the effectiveness of human figure drawings as a clinical instrument for evaluating personality. Unpublished doctoral dissertation, New York Univer., 1952.
- FRENKEL-BRUNSWICK, ELSE. Personality theory and perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality*. New York: Ronald Press, 1951.
- GLOVER, E. Review of Rado, S. *Psychoanalysis of behavior: Collected papers*. *Psychoanal. Quart.*, 1957, **26**, 251-258.
- GOODMAN, H. Self-insight, empathy and perceptual distortion. Unpublished doctoral dissertation, New York Univer., 1952.
- HARROWER, MOLLY R. *Appraising personality: The use of psychological tests in the practice of medicine*. New York: Norton, 1952.
- HERTZ, MARGUERITE. The validity of the Rorschach method. *Amer. J. Orthopsychiat.*, 1941, **11**, 512-519.
- HERTZ, MARGUERITE, & RUBINSTEIN, B. B. A comparison of three "blind" Rorschach analyses. *Amer. J. Orthopsychiat.*, 1939, **9**, 293-314.
- KLOPPER, B., & KELLEY, D. M. *The Rorschach technique*. Yonkers: World Book, 1942.
- KRUGMAN, JUDITH. A clinical validation of the Rorschach with problem children. *Rorschach Res. Exch.*, 1942, **5**, 61-70.
- LITTLE, K. B., & SHNEIDMAN, E. S. The validity of thematic projective technique interpretations. *J. Pers.*, 1955, **23**, 285-294.
- MINTZ, ELIZABETH. Relationships between diagnostic errors and personal anxieties of psychologists. Unpublished doctoral dissertation, New York Univer., 1955.
- MURRAY, H. A. *Thematic Apperception Test manual*. Cambridge: Harvard Univer. Press, 1943.
- NORMAN, R. D. The inter-relationships among acceptance-rejection, self-other identity, insight into self, and realistic perception of others. *J. soc. Psychol.*, 1953, **37**, 205-235.
- PALMER, J. O. A dual approach to Rorschach validation: A methodological study. *Psychol. Monogr.*, 1951, **65**, No. 8 (Whole No. 325).
- PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases*. New York: McGraw-Hill, 1947.
- SAMUELS, H. The validity of personality trait ratings based on projective techniques. *Psychol. Monogr.*, 1952, **66**, No. 5 (Whole No. 337).
- SAXE, E. A quantitative comparison of psychodiagnostic formulations from the TAT and therapeutic contacts. *J. consult. Psychol.*, 1950, **14**, 116-127.
- SCHAFER, R. *Psychoanalytic interpretation in Rorschach testing*. New York: Grune & Stratton, 1954.
- SEARS, R. R. Experimental studies on projection. *J. soc. Psychol.*, 1936, **7**, 151-165.
- SHNEIDMAN, E. *Thematic test analysis*. New York: Grune & Stratton, 1951.
- SIEGEL, MIRIAM. The diagnostic and prognostic validity of the Rorschach test in a child guidance clinic. *Amer. J. Orthopsychiat.*, 1948, **18**, 119-133.
- SILVERMAN, L. A Q sort study of the validity of evaluations made from projective techniques. Unpublished doctoral dissertation, New York Univer., 1958.
- STEPHENSON, W. *The study of behavior: A technique and its methodology*. Chicago: Univer. of Chicago Press, 1953.
- SYMONDS, P. M. A contribution to our knowledge of the validity of the Rorschach. *J. proj. Tech.*, 1955, **19**, 152-162.
- WAEHNER, T. S. Formal criteria for the analysis of children's drawings. *Amer. J. Orthopsychiat.*, 1942, **12**, 95-103.

(Accepted for publication November 5, 1958)

APPENDIX

Q SORTS

I. DEFENSES

1. The patient heavily relies on repression.
2. Intellectualization is prominently used as a defense against sexual needs.
3. Undoing is a frequently used defense mechanism.
4. Isolation of feeling appears to be a major defense.
5. The patient puts great reliance on the defense of avoidance.
6. The patient frequently projects his inadequacies onto others.
7. The patient often alleviates anxiety through aggressive acting out.
8. Overcompensation is a major defense mechanism.
9. The patient often uses depersonalization for defensive purposes.
10. Displacement is one of the most heavily relied upon defenses.
11. The patient often uses the defense of reaction formation against unacceptable passive longings.
12. There is a frequent retreat into fantasy life.
13. The patient has a marked tendency to withdraw from environmental stimulation.
14. The patient frequently feels gay and/or frivolous as a defense against depression.
15. The patient often identifies with the aggressor as a means of warding off anxiety.
16. Somatization is frequently used in an attempt to bind anxiety.
17. The patient uses reaction formation a good deal as a defense against aggressive impulses.
18. The defense of intellectualization as a means of coping with unacceptable dependency needs is prominent.
19. The patient resorts to much sexual acting out as a way of alleviating anxiety.
20. The patient often reacts in a counter-phobic fashion when in a threatening situation.
21. Suppression of impulses frequently takes place.
22. The patient frequently hates those of the opposite sex as a defense against his fear of them.
23. The patient often uses depression as a defense against aggressive feeling.
24. The patient makes frequent use of sublimation.
25. The patient often turns aggressive feeling in on the self.
26. The patient frequently projects his hostile impulses.
27. Denial is used as a major defense mechanism.
28. The patient frequently utilizes regression as a defense maneuver.
29. The patient often projects sexual feeling by seeing others as desirous of him (her).

30. The patient often uses intellectualization as a means of coping with anxiety arousing hostile needs.

II. MOTIVATING NEEDS AND AFFECTS

1. A major source of anxiety for him (her) is fear of separation from maternal figures.
2. The patient suffers much guilt over heterosexual impulses.
3. Competitive strivings with the parent of his (her) own sex play an important motivating role.
4. The fear of losing control over aggressive impulses is present to a significant degree.
5. The patient has prominent fears of being destroyed.
6. A marked need to replace the same sex parent is evident.
7. Fear of helplessness is prominent.
8. Anal sadistic strivings play an important role in the patient's illness.
9. Incestuous strivings toward the opposite sex parent is a strong motivator.
10. The search for an omnipotent father figure plays an important role in motivating the patient.
11. The need to maintain infantile omnipotence is an important motivator.
12. The fear of losing control over heterosexual impulses is present to a significant degree.
13. The fear that his (her) love offerings will be rejected by maternal figures is prominent.
14. Homosexual needs play a prominent motivating role.
15. The patient suffers much guilt over hostile impulses.
16. A search for maternal love is an important motivator for the patient.
17. The patient shows prominent exhibitionistic-voyeuristic needs.
18. A sense of sexual inadequacy frequently motivates the patient.
19. The need to assert his (her) independence is an important motivating factor.
20. The wish for punishment motivates much of the patient's behavior.
21. Feelings of inferiority are a strong motivator.
22. There is a strong fear of castration.
23. Fear of separation from paternal figures is an important motivator.
24. An excessive need for protection and security is present.
25. The fear that his love offerings will be rejected by paternal figures is strong.
26. The patient is strongly motivated by feelings of shame over unacceptable impulses.

27. The patient is much motivated by the need of approval from maternal figures.
28. Strong oral-dependency needs are an important motivator.
29. The patient is much motivated by the need for approval from paternal figures.
30. Feelings of rivalry with siblings play an important role in the patient's illness.

III. CHARACTER TRAITS

1. This is a very ambitious individual.
2. Orderliness is a prominent character trait.
3. He (she) is characterized by much impulsivity.
4. He (she) would be considered a highly self-sufficient person.
5. The patient is a very rigid individual.
6. The patient is much concerned with humanitarian causes.
7. He (she) is notably self-righteous.
8. Evasiveness is an obvious characteristic of this patient.
9. The patient frequently identifies with authority figures.
10. The patient is often not able to complete tasks he begins.
11. This individual is much concerned with physical strength.
12. The patient is highly suspicious.
13. Grandiosity is a noteworthy character trait.
14. This patient has many masculine characteristics (in regard to manner or interests).
15. Ambivalence is a noteworthy character trait.
16. This is a highly egocentric individual.
17. The patient frequently identifies with the underdog.
18. This individual is noticeably feminine in regard to his (her) manner or interests.
19. This is a highly optimistic individual.
20. The patient is very class conscious.
21. The patient is a highly creative person.
22. This is a very moral and/or ethical individual.
23. The patient is characterized by his (her) great concern with wealth and/or power.
24. He (she) is very naive.
25. This is a very parsimonious individual.
26. Lability is an outstanding character trait.
27. He (she) is a very obstinate person.
28. Perseverance characterizes this person to a large degree.
29. The patient is notably haphazard in his (her) approach to things.
30. Much indecision is characteristic of this patient.

IV. DIAGNOSIS AND SYMPTOMS

1. The patient is vulnerable to hysterical phobias.
2. The patient is a passive-aggressive character.
3. The patient is vulnerable to specific obsessions.
4. The patient shows noticeable mood swings between elation and depression.

5. The patient suffers from an obsessive-compulsive character disorder.
6. The patient is vulnerable to anxiety states.
7. The patient is vulnerable to psychotic depression.
8. The patient is vulnerable to psychosomatic symptoms.
9. The patient is vulnerable to simple schizophrenia.
10. The patient suffers from reactive depressions, frequently.
11. The patient is vulnerable to paranoid schizophrenia.
12. The patient is vulnerable to sexual perversion.
13. The patient is suffering from an hysterical character disorder.
14. The patient shows noteworthy psychopathic features.
15. The patient is vulnerable to sexual impotency (frigidity).
16. The patient is suffering from a narcissistic character disorder.
17. The patient is a schizoid character.
18. The patient is vulnerable to hysterical conversion symptoms.
19. The patient is vulnerable to catatonic schizophrenia.
20. The patient is a paranoid character.
21. The patient is vulnerable to compulsive rituals.
22. The patient is a productive or genital character.
23. The patient is a passive-dependent character.
24. The patient suffers from a borderline state utilizing both psychotic and neurotic mechanisms.
25. Organic brain damage is present.
26. The patient is vulnerable to manic states.
27. The patient is vulnerable to drug or alcohol addiction.
28. The patient is vulnerable to a neurotic depression.
29. The patient is vulnerable to amnesia, multiple personality, or other dissociated conditions.
30. The patient is vulnerable to hebephrenic schizophrenia.

V. INTERPERSONAL BEHAVIOR

1. The patient devaluates and mocks those of the opposite sex.
2. The patient is compliant with authority figures.
3. The patient relates to others in an overly intellectual manner.
4. The patient sexualizes his (her) relationships with most persons of the opposite sex.
5. The patient acts in a domineering manner toward those of the same sex.
6. The patient is inclined to play the role of a "clown" when with others.
7. The patient has a Pollyannaish orientation in relating to others.

8. The patient is sadistic in his (her) relationships with those of the opposite sex.
9. The patient withdraws from close contact with those of his (her) sex.
10. The patient generally assumes a passive-compliant role when with peers.
11. The patient acts in a self-assertive manner when with peers.
12. The patient acts toward others in a subtly negativistic fashion.
13. The patient is timid in relating socially to those of the opposite sex.
14. The patient remains inaccessible and aloof from others.
15. The patient withdraws from sexual contact with those of the opposite sex.
16. The patient acts demandingly toward those of the opposite sex.
17. The patient play-acts much of the time.
18. The patient tends to control those of the opposite sex through his (her) weakness.
19. The patient acts in an aggressively grasping attitude toward others.
20. The patient plays a self-effacing role with others.
21. The patient acts competitively toward peers.
22. The patient acts defiantly with authority figures.
23. The patient is pretentious and/or exhibitionistic in relating to others.
24. The patient dependently clings to those of the opposite sex.
25. The patient goes to great lengths to impress others with his (her) juvenile innocence and harmlessness.
26. The patient tends to control those of the same sex through his (her) weakness.
27. The patient tries to domineer those of the opposite sex.
28. The patient acts in an ingratiating manner toward authority figures.
29. The patient is timid in relating socially to those of the same sex.
30. The patient acts competitively toward authority figures.
4. The patient saw the parent of the same sex as too powerful to be opposed.
5. The patient felt that it was dangerous to be assertive except in subterfuge.
6. The patient saw the mother in the role of protector against the patient's father.
7. The patient felt that the father was over-demanding of him (her).
8. The patient perceived the father as passive and submissive.
9. The father was seen as crude and unsympathetic.
10. The patient felt the mother showed love only when her high standards were met.
11. The patient felt the father showed much love for him (her).
12. The patient saw the father as a bully.
13. The patient felt that infantile behavior would be severely punished by one or both parents.
14. The mother was seen as too weak to be of help to the patient.
15. The patient felt it was less threatening to identify with the opposite sex than the same sex parent.
16. The patient felt that the mother was overly punitive.
17. The patient saw the father in the role of protector against the patient's mother.
18. The patient felt that the same sex parent was not strong enough to identify with.
19. The patient felt that his (her) mother was disappointed in him (her) frequently.
20. The mother was perceived as understanding and/or sympathetic.
21. The patient felt that his (her) mother did the thinking for him (her) allowing little autonomy of action.
22. The patient felt that any hostile action or expression would be quickly squelched by one or both parents.
23. The father was perceived as patient and understanding.
24. The patient felt that the mother was aloof from him (her).
25. The father was perceived as overcontrolling.
26. The mother was felt by the patient to be untrustworthy.
27. The patient felt that one or both parents was overpermissive, letting him (her) "get away with murder."
28. The patient perceived the father as unreliable and undependable.
29. The patient felt that the opposite sex parent was sexually seductive toward him (her).
30. The patient perceived the mother as "showering" love and affection on him (her).

VI. INFANCY AND CHILDHOOD AS PERCEIVED BY THE PATIENT

1. The patient felt deprived in his (her) quest for receptive-dependent gratifications.
2. The patient felt that the father was disappointed in him (her) frequently.
3. The patient felt unwanted by the father.





